

عنوان مقاله:

استفاده از معیار شباهت و فاصله در بهینه سازی درخت تصمیم

محل انتشار:

اولین کنفرانس ملی آنالیز داده ها (سال: 1402)

تعداد صفحات اصل مقاله: 8

نویسندگان:

سمیرا اسدزاده - دانشجوی دکتری، گروه کامپیوتر، دانشگاه آزاد اسلامی، واحد شیراز، شیراز، ایران

یاسمن علی اکبریور - دانشجوی دکتری، گروه کامپیوتر، دانشگاه آزاد اسلامی، واحد شیراز، شیراز، ایران

الهام پروین نیا - استادیار، گروه کامپیوتر، دانشگاه آزاد اسلامی، واحد شیراز، شیراز، ایران

خلاصه مقاله:

الگوریتم درخت تصمیم به دلیل دقت بالا، هزینه محاسباتی کم و قابلیت تفسیر بالا به طور گسترده در داده کاوی و یادگیری ماشین استفاده شده است. با این حال، زمانی که با داده های پیوسته سروکار داریم، الگوریتم درخت تصمیم کلاسیک باید با استراتژی گسسته سازی، ویژگی های پیوسته را با ویژگی های گسسته جایگزین کند. گسسته سازی ممکن است باعث از بین رفتن ساختار اطلاعات شود که بر عملکرد طبقه بندی تاثیر می گذارد. در این مقاله، یک شباهت همسایگی جدید تعریف شده. که با ترکیب شباهت با منطق هندسی و شباهت با منطق جبری، چهار نوع جدید از شباهت ها ی همسایگی پیشنهاد شده. و یک الگوریتم درخت تصمیم جدید پیشنهاد شده، که در آن درجه وابستگی ویژگی به عنوان اندازه گیری پارتیشن استفاده شده. نتایج تجربی بر روی ۷ مجموعه داده انتخاب شده از مخزن یادگیری ماشین UCI نشان می دهد که الگوریتم پیشنهاد شده موثر است. میانگین دقت الگوریتم پیشنهادی بیش از ۹۰٪ است که ۱۰٪ بیشتر از الگوریتم های درخت تصمیم گیری کلاسیک است و تعداد گره ها ی برگ کمی افزایش می یابد.

کلمات کلیدی:

بهینه سازی، درخت تصمیم، شباهت، همسایگی

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/1669289>

